

# Automated Mapping of Pharmacy Orders from Two Electronic Health Record Systems to RxNorm within the STRIDE Clinical Data Warehouse

Penni Hernandez, N.D., R.N., Tanya Podchiyska, Susan Weber, Ph.D.,

Todd Ferris, M.D., M.S., Henry Lowe, M.D.

Center for Clinical Informatics, Stanford University, Stanford, CA

## Abstract

*The Stanford Translational Research Integrated Database Environment (STRIDE) clinical data warehouse integrates medication information from two Stanford hospitals that use different drug representation systems. To merge this pharmacy data into a single, standards-based model supporting research we developed an algorithm to map HL7 pharmacy orders to RxNorm concepts. A formal evaluation of this algorithm on 1.5 million pharmacy orders showed that the system could accurately assign pharmacy orders in over 96% of cases. This paper describes the algorithm and discusses some of the causes of failures in mapping to RxNorm.*

## Introduction

The Stanford Translational Research Integrated Database Environment (STRIDE)<sup>1</sup> is an informatics research and development project at Stanford University Medical Center (SUMC) to create a standards-based informatics platform supporting clinical and translational research (CTR). STRIDE receives clinical data, for research use, via HL7 messages from SUMC information systems supporting patient care at Lucile Packard Children's Hospital at Stanford (LPCH) and Stanford Hospital & Clinics (SHC). This data is integrated into the STRIDE Clinical Data Warehouse (CDW), an Oracle-based system that uses a data model based on the HL7 Version 3 Reference Information Model (RIM)<sup>2</sup>. STRIDE supports integrated access to clinical data, for research purposes, from the pediatric and adult patient populations at SUMC. A Java application, called the STRIDE Anonymous Patient Cohort Identification Tool, gives Stanford researchers the ability to identify research patient cohorts in the CDW, using a variety of clinical criteria, without exposing protected health information (PHI).

Medication information is an important data type for CTR. Accurate, standards-based, representation of medications assures a common understanding of the data, which facilitates retrieval, analysis, and sharing of pharmacy data for CTR. However, many clinical

and pharmacy systems use drug information databases from commercial vendors, which may use different proprietary identifiers, naming conventions and drug models. This is the case at SUMC, where LPCH and SHC operate two separate EHR systems that use different commercial drug databases. Thus, even though SUMC hospitals are cooperating to share content with STRIDE, their data are incompatible. To support integrated representation and searching of pharmacy data across both SUMC hospitals, STRIDE needed a standards-based drug representation model within its CDW.

The National Library of Medicine (NLM) and the Food and Drug Administration (FDA) set out to standardize drug information identifiers to support interoperability by creating RxNorm<sup>3</sup>, a free, robust and current drug representation system, which is updated weekly. RxNorm allows navigation between ingredients, generic drug names, brand names, and National Drug Codes (NDC) identifiers through the use of defined relationships. RxNorm is one of the source vocabularies of NLM's Unified Medical Language System (UMLS). It provides a unified drug representation model and maintains a mapping between different proprietary drug identifiers. The major drug information vendors submit some level of their terminologies to the UMLS for mapping within RxNorm.

This paper describes the use of RxNorm as a standards-based drug representation model within STRIDE. RxNorm and its built-in relationships were leveraged to provide mapping between pharmacy data from two SUMC EHR systems employing different proprietary drug vendor information systems. In particular, we are interested in the following outcomes: (1) RxNorm coverage for the drug concepts derived from two sources of SUMC pharmacy orders; (2) utilization of the linkages within RxNorm, particularly those linking brand names to generic ingredients (3) characterization of the pharmacy message that could not be automatically mapped to RxNorm and (4) mapping from RxNorm concepts to the SNOMED-CT substance hierarchy.

The approach of using algorithms to map biomedical concepts to standardized terminologies, followed by manual review of the mapping results by medical domain experts, is well-documented<sup>4</sup>. Alternative approaches to integration of drug terminologies include the use of ontologies<sup>5</sup>. RxNorm has been used to extract drug names from narrative text clinical documents<sup>6</sup> and for computable exchange of drug allergy information between the Department of Veterans Affairs (VA) and the Department of Defense (DoD)<sup>7</sup>. RxNorm was selected as a Consolidated Health Informatics (CHI) designated standard for Trade names and Drug Names. RxNorm and the VA National Drug File Reference Terminology's (NDF-RT)<sup>8</sup> are the recommended standards for representing drug names and drug classification.

Given RxNorm's emerging role as a national standard, its use within STRIDE was felt to offer a scalable strategy for representing drug orders obtained from different EHR systems using different drug vendor information models. This approach may be of interest to others who have to merge pharmacy data from multiple clinical systems into a common standards based representational framework.

## Methods

STRIDE receives several types of HL7 messages containing drug information from both SUMC hospitals. While the Pharmacy Order (RDE), the Pharmacy Dispense (RDS) and the Detailed Financial Transaction (DTF) messages all contain data about drugs ordered, each has limitations that needed to be considered. The Pharmacy Dispense messages are used by drug dispensing devices like Pyxis, but only about 20% of medications ordered at LPCH are dispensed through a device, while 80% are custom compounded. For adult pharmacy orders at SHC, the opposite is true, with approximately 80% dispensed through a device, while only 20% are custom compounded. The DFT messages did not contain a robust description of the medication form and dosing.

HL7 v2.3 SUMC Pharmacy Order (RDE) messages were selected as the initial source of pharmacy information to be loaded into the STRIDE CDW. The goal was to achieve a complete mapping of each hospital pharmacy order to RxNorm Ingredient (IN) concepts. An algorithm was developed in Oracle PL/SQL to match data received in the HL7-based Pharmacy Orders to RxNorm atoms of type IN. The RxNorm IN name and RXCUI were used as the target terminology mapping level.

Each RDE message contains three HL7 v2.3 segments of interest:

1. The Pharmacy Encoded Order (RXE) segment contains data used to manage drug ordering within the Stanford hospitals. The drug order information is located within the "Give Code" data field of the RXE segment. This data field provides three components as data sources for the algorithm. The first component is the "Give ID" which is the assigned local ID code for the order. The second component conveys the drug name, form and strength and will be referred to as "Give Text." The third relevant field component delivers an alternate drug order representation without the drug strength and will be called "Give Alt Text". Due to differences of implementation between the two hospitals, the "Dispense Package Method" data field of the RXE segment was also used to provide supplementary information on suggested brand names.

The "Give ID" and "Give Text" were extracted and stored unchanged in the STRIDE RIM-based CDW data model as the definitive reference data for each RXE segment.

2. The Pharmacy Component (RXC) segment encodes data, similar to the RXE segment, on the specific components of the order.

3. The Pharmacy Route (RXR) segment encodes the route of administration.

The combination of RXE, RXC and RXR segments of RDE messages fully defined the drug order. An example of these segments is given below:

```
RXE|RXCUST_IV^^PYXIS^^oxytocin additive 20
units + Lactated Ringers Injection 1000 mL
RXC|A|OXYTOB201^Oxytocin
RXR|IV^IV
```

Our goal was to map every unique drug order text string from the HL7 messages to its corresponding ingredients and route of administration. The "Give Text" tended to be more precise from the standpoint of listing ingredients, so we used it as our preferred data source and we utilized the "Give Alt Text" only when we were unable to extract the expected list of ingredients, based on the count of ingredient separators, from the medication order in the "Give Text".

## Algorithm steps

1. Select all unique combinations of Give ID, Give Text, and Give Alt Text from each pharmacy order.
2. Use the Give Text for a given pharmacy order as the input string.

3. Compare only the alphanumeric characters of the input string to the STR column in the RxNorm concept table RXNCONSO, ignoring case.

4. If an exact match for the alphanumeric component is found, which has an RXCUI for a non-suppressed concept in the RxNorm vocabulary with TTY in ('IN', 'BN', 'GPCK', 'BPCK', 'SCD', 'SCDC', 'SCDF', 'SBD', 'SBDC', 'SBDF') then that concept becomes the starting point for mining INs for the pharmacy order. Proceed to step 7.

5. If more than one match is found for the string, skip to step 10.

6. If no matches are found in RxNorm and the input string is longer than one word, then remove the last word from the input string and use that new string as input to step 3. If there is no whitespace left and no match has been found, but a dash is present, then we use the substring before the dash as input to step 3.

7. Check the TTY of the concept, where:

BN = Brand Name, BPCK = Branded Pack, GPCK = Generic Pack, IN = Ingredient, PIN = Precise Ingredient, SBD = Semantic Branded Drug, SBDC = Semantic Branded Drug Component, SBDF = Semantic Branded Drug Form, SCD = Semantic Clinical Drug, SCDC = Semantic Clinical Drug Component, SCDF = Semantic Clinical Drug Form

7.a - If TTY='BPCK', save concept, find its SBDs using the 'contained\_in' relationship in the RXNREL table, and use each associated SBD concept as input to step 7.c

7.b - If TTY='GPCK', save concept, find its SCDs using the 'contained\_in' relationship in the RXNREL table, and use each associated SCD concept as input to step 7.d

7.c - If TTY in ('SBD', 'SBDC', 'SBDF'), save concept, get its BNs using the 'ingredient\_of' relationship in the RXNREL table, and use each resulting BN concept as input to step 7.f

7.d - If TTY = 'SCD', save concept, find its SCDCs using the 'constitutes' relationship in the RXNREL table, and use each associated SCDC concept as input to step 7.e

7.e - If TTY in ('SCDC', 'SCDF'), save concept, find its INs using the 'ingredient\_of' relationship in the RXNREL table, and use each resulting IN concept as input to step 7.g

7.f - If TTY = 'BN', save concept, find its INs using the 'has\_tradename' relationship in the RXNREL

table, and use each resulting IN concept as input to step 7.g

7.g - If TTY = 'IN', check whether the mapped concept is not a PIN (a salt form, an isomer, or some other lexical variant) for a clinically significant IN concept, by using the 'has\_form' relationship in RXNREL table. If no results returned for relationship among ingredients, then go to step 8 with current RXCUI. If clinically significant ingredient(s) are present for RXCUI, save those in step 8.

8. Save IN's RXCUI, STR, TTY, if not already saved for current pharmacy order.

9. Check for presence of a SNOMED vocabulary atom of TTY='FN' (Fully Specified Name) for saved RxNorm IN RXCUI. Save both product and substance SNOMED codes as possible future seeds into SNOMED for traversal to substance class.

10. If multiple ingredients are present in the input text as indicated by the presence of any of the characters in the following regular expression range [-&,^+()] and the count of mined ingredients is less than the count of separators plus one, then split the array on the separators and use each string of the array as an input to step 3.

11. If the number of mapped ingredients is still less than the count of separators plus one, then use the Give Alt Text, Component Alt Text, or supplementary field as input to step 3, and flag all resulting concepts as derived from Give Alt Text, also flag the pharmacy order for manual review.

12. If separators defined in the range [-&,^+()] are present in the Give Alt Text, then split the alt text into a string array based on the delimiters above and feed each array element as an input to step 3.

The mapping algorithm was evaluated using 15 weeks of HL7 RDE pharmacy order messages from both Stanford hospitals. This test set contained 1,203,962 RXE|RXC segments with 2,346 unique pairs from SHC and 390,792 RXE|RXC segments with 7190 unique pairs from LPCH.

Clinician experts reviewed and validated all of the RxNorm concepts assigned by the algorithm to the unique Pharmacy Orders in the test set. The mapping algorithm included detailed logging to assist with identification of the assignment origin (e.g. ingredients derived from Brand Name matches). For each mapping run a table of RxNorm IN was generated for each unique Give ID, Give Text and Give Alt Text along with all flags from processing.

The RxNorm concepts assigned by the algorithm were manually categorized as follows:

- True Positive – Algorithm accurately mapped all ingredients in the pharmacy order to the appropriate RxNorm concepts.
- True Negative – Algorithm correctly determined that RxNorm did not contain any appropriate concepts for the ingredients in the pharmacy order being processed.
- False Positive – Algorithm mapped one or more of the ingredients in the pharmacy order to incorrect RxNorm concepts.
- False Negative – Algorithm failed to map all ingredients in the pharmacy order to appropriate existing RxNorm concepts.

The expert reviewers based the gold standard for the mapping on manual evaluation of the assigned RxNorm concepts. When the algorithm failed to map an HL7 pharmacy order message to RxNorm, a clinician attempted to manually map the ingredients to RxNorm using NLM's RxNav interface<sup>9</sup>.

## Results

Combined results of mapping messages from both hospitals are in table 1. Separate results for LPCH and SHC messages are in tables 2 and 3 respectively.

	Total Number	%
True Positive	8895	93.28
True Negative	316	3.31
False Positive	305	3.20
False Negative	20	0.21

**Table 1.** All Unique Pharmacy Orders

	Total Number	%
True Positive	6665	92.70
True Negative	270	3.76
False Positive	240	3.34
False Negative	15	0.21

**Table 2.** Unique Pharmacy Orders From LPCH

	Total Number	%
True Positive	2230	95.06
True Negative	46	1.96
False Positive	65	2.77
False Negative	5	0.21

**Table 3.** Unique Pharmacy Orders From SHC

## Discussion

The algorithm correctly mapped 93.28% of pharmacy messages to RxNorm (True Positives). It also correctly determined that no appropriate mapping to RxNorm was possible for 3.31% of messages (True Negatives). Thus the algorithm correctly assigned 96.59% of pharmacy messages. We examined the 316 True Negatives and categorize them in table 4.

Category	N
Text fragments in message (e.g. 'Ferric')	83
Vitamins (e.g. 'Prenatal Multivitamin')	42
Peritoneal dialysis solutions	38
Certain vaccines	36
Over the counter items (e.g. 'Eucerin')	28
Custom mixture (e.g. 'GI Cocktail')	23
Dummy orders for workflow management	22
Abbreviations not present in RxNorm	17
Investigational Drugs not in RxNorm	12
Brand names not in RxNorm	7
Non-drug substance (e.g. Avitenet)	5
Typographical errors in message	2
Device (e.g. space device for inhaler)	1

**Table 4.** Categories of True Negatives

There were a variety of reasons why the algorithm incorrectly assigned RxNorm concepts to pharmacy messages (false positives). One major cause was that

the algorithm used an exhaustive set of the text delimiters possible in the pharmacy orders. This improved the mapping sensitivity but reduced the level of specificity. For example the algorithm parsed “Epinephrine, racemic” twice: one parse gleaned “Racepinephrine” and the second “Epinephrine”. The presence of a comma in the pharmacy order was interpreted by the algorithm as an indicator of the presence of two ingredients. Another cause of false positives was the use of ingredient descriptors within the pharmacy order separated from the medication name by a defined delimiter. For example, the character string “/INH”, where the algorithm considered INH a potential ingredient and matched it to the drug Isoniazide, instead of recognizing it as shorthand for “inhaler”.

## Conclusion

The algorithm we developed uses a number of lexical methods to automate the mapping of drug terminology from two Electronic Health Record Systems that use different drug representation systems to RxNorm within a clinical data warehouse. The version of the algorithm evaluated in this paper correctly mapped approximately 93% of SUMC pharmacy orders to RxNorm concepts. No suitable RxNorm concept could be found (algorithmically or manually) for about 3% of the pharmacy orders. We have described the general categories of these failures. For the approximately 4% of pharmacy orders where the algorithm failed to map to an existing RxNorm concept or mapped to the wrong concept, we have identified the source of these failures. In some cases the complexity of the data within an order class will require manual mapping to RxNorm. It is important to note that manually mapped orders, once verified, do not need to be manually mapped again. Inbound HL7 pharmacy messages that do not map algorithmically in STRIDE are forwarded to a human expert for review and a mapping table. After the initial phase of manually mapping non-matching orders, the number of additional orders requiring manual mapping will be quite small. Future plans for the project include migration to other CHI/HITSP drug standards like NDF-RT chemical drug classes and extending the algorithm to handle allergy information. We are also interested in assessing how RxTerms<sup>10</sup>, a drug interface terminology being developed by NLM, might be useful in this work.

An additional benefit of this mapping project that we have yet to evaluate is the ability to derive SNOMED-CT drug classes for ingredients mapped to RxNorm. Each SUMC hospital uses a different proprietary drug classification. In order to query the unified drug data within STRIDE by drug class we needed a drug classification content set mapped to

RxNorm ingredients. One classification available is the SNOMED-CT drug classification tree. Once the RxNorm ingredient RXCUI had been retrieved, that identifier is used to find the corresponding concept ID in the SNOMED-CT substance hierarchy. The defining SNOMED-CT “is-a” relationships are traversed to retrieve the classification. The preferred approach for the future would be to map to the NDF-RT drug classification. This set has been designated by the government as the standard drug classification to be used with RxNorm. However, a publicly available mapping between NDF-RT and RxNorm does not currently exist. Version 2008AA\_081001F of RxNorm was used in this evaluation.

## References

- 1 <http://stride.stanford.edu>
- 2 Schadow G, Mead CN, Walker DM. The HL7 reference information model under scrutiny. *Stud Health Technol Inform* 2006;124:151-6.
- 3 <http://www.nlm.nih.gov/research/umls/rxnorm/>
- 4 Barrow R, Cimino JJ, Clayton, P. Mapping Clinical Useful Terminology to a Controlled Medical Vocabulary. *AMIA Symp*. 1994:211-215.
- 5 Wroe CJ, Cimino JJ, Rector AL. Integrating existing drug formulation terminologies into an HL7 standard classification using opengalen. *Proc AMIA Symp* 2001:766-70.
- 6 Levin MA, Krol M, Doshi AM, Reich DL. Extraction and mapping of drug names from free text to a standardized nomenclature. *AMIA Annu Symp Proc* 2007:438-42.
- 7 Warnekar PP, Bouhaddou O, Parrish F, Do N, Kilbourne J, Brown SH, Lincoln MJ. Use of rxnorm to exchange codified drug allergy information between department of veterans affairs (VA) and department of defense (dod). *AMIA Annu Symp Proc* 2007:781-5.
- 8 Brown SH, Elkin PL, Rosenbloom ST, Husser C, Bauer BA, Lincoln MJ, et al. VA national drug file reference terminology: A cross-institutional content coverage study. *Stud Health Technol Inform* 2004;107(Pt 1):477-81.
- 9 Zeng K, Bodenreider O, Kilbourne J, Nelson S. Rxnav: A web service for standard drug information. *AMIA Annu Symp Proc* 2006:1156.
- 10 Fung KW, McDonald C, Bray BE. Rxterms a drug interface terminology derived from rxnorm. *AMIA Annu Symp Proc* 2008:227-31.